Обзор методов идентификации и аутентификации пользователей по голосу

Мамырбаев Оркен Жумажанович

ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институты

Кыдырбекова Айзат Сиязбековна

әл-Фараби атындағы Қазақ Ұлттық университеті

Мұса Тұрдалыұлы

ҚР БҒМ ҒК Ақпараттық және есептеуіш технологиялар институты

Аннотация. В этой статье описывается методы, который сочетает в себе обучение модели со многими состояниями (multicondition model training) и теории неполных данных (missing-feature theory) к модели шума с неизвестными спектрально-временными характеристики.

**Ключевые слова:** биометрия, биометрическая технология, голосовая верификация, аутентификация, методы распознавания, голосовая биометрия, идентификация по голосу.

Abstract. In this article we describe methods, which involves the study of multicondition model training and the theory of missing-feature theory. The model is characterized by spectral-time characteristic.

Key words: biometrics, biometric technology, voice authentication, methods of identification, voice biometrics, Identification by voice, voice recognition.

Аңдатпа. Мақалада дауысты идентификациялаудың және аутентификациялаудың әдіс-тәсілдері сипатталады, ол көптеген жағдайларда (multicondition model training) және толық емес деректердің теориялық (missing-feature theory) модельдердің шашыраңқы спектральды-уақыттық сипаттамалары қарастырылған.

Түйінді сөздер: биометрия, биометриялық технологиялар, дауыстық аутентификация, верификация, дауыстық биометрия, дауыс бойынша идентификациялау.

Введение

Одной из распространенных мер защиты информации является разграничение доступа. Разграничение доступа включает в себя аутентификацию – процедуру проверки подлинности, и идентификацию – присвоение идентификатора и сравнение предъявленного идентификатора с перечнем идентификаторов [1]. Существует более простая задача – верификации, подразумевающая сравнение предъявляемых признаков, заявленных как соответствующие известному объекту, с известными признаками того же объекта. В современном мире широко развиваются речевые технологии различного применения – синтеза, идентификации, распознавания речи. Последние две технологии позволяют выделить из речи информацию различного типа – «кто» говорит и «что» именно говорится. Предлагается использовать популярный в текстонезависимой идентификации диктора метод выделения i-векторов для решения задачи идентификации говорящего.

Среди множества различных методов разграничения доступа особое место занимают биометрические технологии. Растущий интерес к биометрическим технологиям связан в основном с удобством их применения. Часто под биометрическими технологиями, в контексте разграничения доступа, понимают методы биометрической идентификации. Задача идентификации подразумевает сравнение признаков одного объекта с признаками других известных объектов.

Ведущую роль среди биометрических технологий занимают технологии, основанные на распознавании отпечатков пальцев, лица, радужной оболочки глаз и голоса. Для определения эффективности технологий оценивают такие критерии, как надежность, устойчивость к подделке, устойчивость к окружающей среде (помехоустойчивость), стабильность признака от времени, скорость, стоимость и удобство применения.

Голосовая идентификация обладает такими достоинствами, как удобство применения и невысокая стоимость. Недостатком такой идентификации является низкая надёжность. Одним из перспективных путей повышения надёжности голосовой идентификация является привлечение характеристик динамики подсознательных движений, что активно используется при идентификации по подписи [2]. С другой стороны, есть области применения, в которых голосовая идентификация является наиболее удобной, например, удалённый доступ к телекоммуникационным каналам связи по анализу голосовых данных.

Принципиальный недостаток всех методов биометрии, кроме речевого, состоит в постоянстве используемого биометрического кода, т.к. отпечатки пальцев или ладоней, рисунок радужной оболочки и черты лица неизменны для индивидуума. Этот недостаток препятствует применению этих методов в случаях, требующих особо высокой надежности идентификации личности, поскольку неизменный биометрический код может быть считан путем злонамеренного вторжения в программу распознавания.

В отличие от биометрии по фиксированным параметрам, верификация по голосу обладает практически неограниченным потенциалом для снижения ошибки за счет использования все более длинных речевых сообщений. Верификации по голосу может использоваться в темноте, на расстоянии, в частности, по стандартному телефонному каналу, в условиях, когда невозможно получить изображение лица.

Примеры конкретных применений верификации диктора охватывают широкий спектр приложений:

- распоряжение финансовыми процессами по электронным или телефонным каналам (управление банковским счетом, электронная коммерция, подтверждение права пользования кредитной картой);

- разрешение на смену пароля или PIN-кода;

- доступ к компьютеру или отдельным программам компьютера (вход в Интернет, доступ к конфиденциальным документам, базам данных и т.д.);

- разрешение на вход в помещение, открывание сейфа;

- управление механизмами и системами (например, запуск двигателя автомобиля);

- мониторинг того, кто, когда и к каким компьютерным ресурсам имел доступ.

В современных системах голосовой идентификации для повышения надежности применяют тексто-зависимую идентификацию, например, произнесение парольной фразы, которая каждый раз генерируется случайным образом. Использование индивидуальных признаков и совпадение сгенерированной и распознанной парольных фраз повышает надежность. Тексто-независимая идентификация подразумевает использование только индивидуальных признаков.

Немаловажной характеристикой системы голосовой идентификацииявляется скорость (быстродействие) определения личности**.** Быстродействие особенно существенно для приложений, обрабатывающих большие базы голосовых данных и работающих в реальном режиме времени. Повышение быстродействия может быть достигнуто за счет использования новых быстрых алгоритмов обработки данных. Таким образом, голосовая идентификация личности, несмотря на указанные в настоящей работе недостатки, в определенных условиях обладает существенными преимуществами, которые необходимо развивать. В задаче голосовой идентификации применяют различные математические, алгоритмические, технические методы, начиная с этапа записи голоса и заканчивая этапом классификации. Практически каждая система идентификации содержит четыре основных этапа: получение сигнала, предварительная обработка сигнала, извлечение признаков и классификация признаков.

Этап получения сигнала. Метод получения или записи голосового сигнала, в большинстве случаев, представляет собой запись сигнала с помощью микрофона и представление сигнала в цифровом виде с помощью аналого-цифрового преобразователя. В качестве аналого-цифрового преобразователя обычно используют звуковую карту персонального компьютера либо цифровой диктофон. Цифровые данные кодируются импульсно-кодовой модуляцией и помещаются в формат файла-контейнера (Waveform Audio File Format) для хранения записи оцифрованного аудиопотока. Параметры звуковой записи обычно выбирают следующие: битность отсчета – 16 бит, частота дискретизации – 22500 Гц. Так как современные цифровые мобильные устройства обычно имеют встроенный микрофон и производительные аппаратные средства, то создание система аутентификации по голосу с привлечением более затратных по вычислениям методов вполне решаемая задача для мобильных платформ. Тем не менее, обеспечить минимальные вычислительные затраты при сохранении точности, помехоустойчивости к различным видам помех и достаточную надежность при распространенных аппаратных средствах все же необходимо.

Этап предварительной обработки. Полученные цифровые сигналы, как и аналоговые, содержат в себе некоторую долю искажений и помех. Под искажениями понимаются искажения речеобразующего тракта (например, болезнь горла) и речепередающего канала (например, искажения телефонного канала).

Этап предварительной обработки. Полученные цифровые сигналы, как и аналоговые, содержат в себе некоторую долю искажений и помех. Под искажениями понимаются искажения речеобразующего тракта (например, болезнь горла) и речепередающего канала (например, искажения телефонного канала).

Этап извлечение признаков. Извлечение признаков обычно происходит с помощью Фурье-преобразования, вейвлет- преобразований, линейного предсказания и других. Коэффициенты преобразований выступают в качестве признаков. В настоящее время точно не определены голосовые признаки, по которым можно однозначно идентифицировать личность человека. Выбор признаков влияет также на надежность идентификации. Существуют, методы которые описывают интегральные характеристики человеческого голоса и служат для извлечения тонов, динамики речи, просодических характеристик. Такими методами являются преобразование Фурье (амплитудно-частотное распределение), кепстральное преобразование (амплитудно-временное распределение), преобразование линейного предсказания (амплитудно-частотное распределение). Существуют также формантные методы и методы выделения фонем.

Этап классификации признаков. В этот этап входит применение математических методов классификации, с помощью которых осуществляется принятие решения, а также расчет ошибок классификации.

Системы распознавания речи строятся на основе принципов признания форм распознавания. Методы и алгоритмы, которые использовались до сих пор, могут быть разделены на следующие большие классы:

Классификация методов распознавания речи на основе сравнения с эталоном.

- Динамическое программирование — временные динамические алгоритмы (Dynamic Time Warping).

- Контекстно-зависимая классификация. При её реализации из потока речи выделяются отдельные лексические элементы — фонемы и аллофоны, которые затем объединяются в слоги и морфемы.

- Методы дискриминантного анализа, основанные на Байесовской дискриминации (Bayesian discrimination);

- Скрытые Марковские модели (Hidden Markov Model);

- Нейронные сети (Neural networks).

Метод динамического программирования. Dynamic Time Warping (DTW) – метод динамического программирования, позволяющий найти близость между двумя последовательностями измерений за некоторый промежуток времени.

Наиболее распространенными из них являются модели гауссовых смесей и скрытые марковские модели. Однако другие модели, например, многослойные персептроны или машина опорных векторов, также успешно используются в данной задаче. Кроме того, в последнее время наблюдается тенденция использования комбинаций нескольких моделей.

За последние годы были достигнуты значительные улучшения в проверке тексто-независимую идентификацию. JFA, как метод компенсации, был более эффективным, чем другие традиционные методы, основанные на GMM-UBM. Самый последний метод использования вероятностного линейного дискриминантного анализа (PLDA) - это современный метод. В последние несколько лет многие из вновь предложенных методов были добавлены и к проверке текстозависимого докладчика.

В системах ЦРТ поиск может осуществляться с помощью метода анализа статистик основного тона (PS), спектрально-формантного (SF) метода, GMM–SVM-метода [3, 4, 5]. Наиболее распространенным является GMM–SVM-метод (рис. 1)[9].

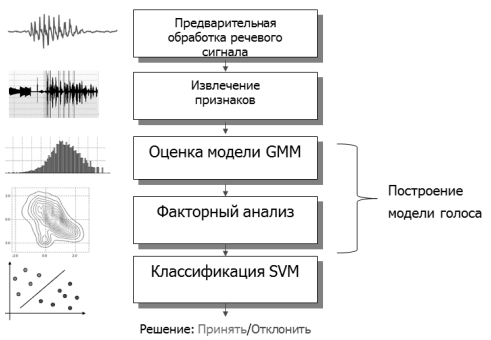


Рис. 1. Схема обработки речевого сигнала GMM–SVM-методом

В [6], помимо введения текстозависимого набора данных RSR2015, оценивается несколько методов. Он отмечает, что отсутствие достаточных данных делает невозможным извлечение i-вектора. Поэтому они используют экстрактор, обученный с использованием набора данных телефонии NIST. Во всех случаях методы, основанные на i-векторе, приводят к снижению точности.

В анализе общего фактора (JFA) [7] они используют Универсальную фоновую модель (UBM), используют набор данных RS42015 и демонстрируют, что адаптация этой модели к каждой фразе дает наилучшие результаты. Напротив, авторы [8] используют фразе-зависимые преобразования PLDA вместо фразе-зависимых UBM и показывают эмпирически.

Идентификация личности по голосу в настоя­щее время реализуется на базе целого ряда принци­пов:

1) используя заданный текст — на основе анали­за и сравнения, начитанных человеком определен­ных слов или фраз, с эталонными;

2) с зависимостью от текста — на основе срав­нения произнесенных фраз с произношением ранее записанного текста;

3) прямая обработка речи — обработка любой фразы говорящего, выделение основных ее параме­тров и характеристик и сравнение их показателей с эталонными;

4) диалогового принципа — классическое рас­познавание речи во время диалога (при общении по телефону и т.д.).

Несмотря на то, что методы идентификации во многом отличаются, в целом можно выделить следующие основные этапы, присущие каждому из рассматриваемых методов:

1. Извлечение признаков из входного речевого сигнала.

2. Построение модели (шаблона) пользователя на основе полученных на предыдущем шаге векторов признаков.

Процесс определения пользователя, зарегистрированного в системе, по входному речевому сигналу во всех рассматриваемых методах состоит в поиске наиболее подходящей сохраненной модели на основе каких-либо критериев.

Некоторые методы предлагают построение модели идентификации: скрытая Марковская модель, Гауссовы смеси. Существуют методы идентификации:

- основанные на классификации по определенным параметрам: метод динамического программирования, векторное квантование.

- для извлечения входных параметров (спек­тральный анализ, преобразование Фурье, линейное предсказание речи, кепстральный анализ) и их рас­познавания. Использование спектрального анализа и преобразования Фурье позволяет полу­чить массив данных, характеризующих спектры речевых фрагментов. Уменьшение объема анали­зируемых данных позволяет достичь использова­ния коэффициентов линейного предсказания речи, кепстральных коэффициентов и мел-частотных кепстральных коэффициенты (MFCC). Подобные единицы измерения активно применяют при реше­нии задач распознавания, так как их использование позволяет приблизиться к механизмам человече­ского восприятия, которое нашло широкое приме­нение в биометрических системах идентификации человека.

**методы распознания**

**сложные**

**простые**

**без обучения**

**с обучением**

**с самообучением**

**структурные**

**комбинированные**

**нейронные сети**

**логические**

**вероятностные**

**детерминированные**

**- метод Байеса;**

**- метод максимума правдоподобия;**

**- линейный дискриминантный анализ;**

**- аппроксимационный метод оценки распределений по выборке;**

**- методы разложения по базисным функциям;**

**- таксономия;**

**- компонентный и факторный анализ**

**1. Сеть прямого распространения:**

**- однослойный перцептрон; - сеть радиальных базисных функций;**

**- многослойный перцептрон**

**2. Рекурентные сети:**

**- соревновательные сети;**

**- сеть Кохенена;**

**- сеть Хопфилда;**

**- модули ANT**

**- методы построения этолонов**

**-метод дробящихся этолонов**

**-линейные решающие правила**

**-метод ближайших соседей**

**- метод потенциальных функций;**

**- структурные (лингвистические) методы;**

**- кластерный анализ**

Рис. 2 Классификация методов распознавания

Существенной проблемой для систем голосовой идентификации, основанных на вышеперечисленных методах, является сильное воздействие внешнего окружающего шума на исходные голосовые записи, из которых выделяются информативные признаки. Обусловленное шумом искажение этих признаков вызывает высокий уровень ошибок идентификации.

Заключение.

Разработка новых методов и средств для предотвращения несанкционированного доступа к информационным ресурсам посторонних лиц требует развития систем и процедур идентификации личности. Особое место занимают системы и способы биометрической идентификации, обладающие целым рядом достоинств и преимуществ. Биометрические показатели, основанные на уникальных индивидуальных особенностях человека, наиболее надежно характеризуют отдельного пользователя. Они достаточно легко подвергаются автоматизированной обработке и хранению в базах данных, что обеспечивает удобство в использовании и относительно небольшие затраты на применение. Среди наиболее перспективных систем биометрической идентификации, основанных на физиологических особенностях человека, большую роль играет голосовая идентификация. Это обусловлено целым рядом преимуществ: акустический амплитудно- частотный спектр речевых сигналов, уникальный для каждого человека, достаточно легко записывается, эти записи удобно хранить в компьютерных базах данных, подвергая при необходимости анализу и обработке. В даннойстатье рассмотрены различные математические модели,использующиеся для построения систем распознавания речи (таких какСкрытые Марковские Модели), особое внимание уделяется методам,применяемым для разработки системы распознавания речи, точностьидентификации которых не зависит отдиктора. Дляреализации этогорассматривается возможность предварительной идентификации диктора,атакже способ построения дикторонезависимых признаков для описанияречевого сигнала, опирающийся на психоакустическую модель восприятиячеловеком речевого сообщения.Проанализированыматематические модели, применяющиесядля построения систем распознавания речи.

Основные результаты работы заключаются в следующем:

1. Проведён анализ существующего состояния в сфере распознавания пользователя по голосу.

2. Рассмотрены методы построения дикторонезависимых систем распознавания речи.

**Использованные литературы**

1. Руководящий документ Гостехкомиссии России. «Защита от несанкционированного доступа к информации. Термины и определения». – М. : ГТК РФ, 1992. – 13 с.

2. Иванов, А.И. Биометрическая идентификация личности по динамике подсознательных движений / А.И. Иванов. – Пенза : Изд-во Пенз.гос.ун-та, 2000. – 187 с.

3. Идентификация дикторов на основе сравнения статистик основного тона го- лоса / С.Л. Коваль, П.В. Лабутин, Е.В. Малая и др. // Информатизация и ин- формационная безопасность правоохранительных органов. Тр. XV Межд. науч. конф. М.: Академия управления МВД России, 2006. С. 324–327.

4. K o v a l S. L. Formants Matching as a Robust Method for Forensic Speaker Identification // SPECOM'2006. Proc. XI Intern. Conf. «Speech and Computer». St. Petersburg, 2006. P. 125–128.

5. Капустин А. И., Симончик К. К. Система верификации дикторов по голосу на основе использования СГР–SVM подхода // DSPA–2010. Тр. 12-й Межд. конф. «Цифровая обработка сигналов и ее применение». М., 2010. Т. 1. С. 207–210.

6. A. Larcher, K. A. Lee, B. Ma, and H. Li, "Text-dependent speaker verification: Classifiers, databases and RSR2015," Speech Communication, vol. 60, pp. 56-77, 2014.

7. P. Kenny, T. Stafylakis, J. Alam, P. Ouellet, and M. Kockmann, "Joint Factor Analysis for Text-Dependent Speaker Verification," in Proc. Odyssey Speaker and Language Recognition Workshop, Joensuu, Finland, 2014.

8. T. Stafylakis, P. Kenny, P. Ouellet, J. Perez, M. Kockmann, and P. Dumouchel, "I-Vector/PLDA Variants for Text-Dependent Speaker Recognition,". http://www.crim.ca/perso/patrick.kenny.

9. Ю. Н . Матвеев Технологии биометрической идентификации личности по голосу и другим модальностям. ВестникМГТУим. Н.Э. Баумана. Сер. “Приборостроение”. 2012